COMPARATIVE STUDIES ON THE CHEMISTRY AND CHAIN STRUCTURE OF COLLAGEN

3543

E. Kulonen and J. Pikkarainen

There is no reason why collagen should be an exception from the evolution of proteins. However, the number of the acceptable mutations is much more restricted by the rigid secondary structure than in globular proteins. As a protein collagen is "young" when compared to most of the intracellular proteins, which also has limited the possible variation.

We wish to summarize our work on the comparative aspects on the chemistry of collagen (Pikkarainen and Kulonen, 1965; Pikkarainen, Rantanen and Kulonen, 1966; Pikkarainen and Kulonen, 1967; Pikkarainen, 1968; Pikkarainen et al., 1968). It was based on and partly inspired by previous knowledge on the composition of collagen (Eastoe and Leach, 1958; Gross, 1963; Eastoe, 1968), which has been related especially with the thermal stability of collagen (Gustavson, 1956; Josse and Harrington, 1964; Harrington and Rao, 1967; von Hippel, 1967).

It is generally accepted that every third residue in collagen must be glycine and that the amount of imino acid residues is correlated rather well with the thermal stability of collagen, assessed either as the denaturation or as the shrinking temperature. Thermal stability is related to the body temperature of the species both in the vertebrates and invertebrates (Nordwig, 1968). There is a certain lower limit in the content of imino acids, necessary to keep collagen as a helix at 0°C (von Hippel and Wong, 1963). The lowest imino acid content and denaturation temperatures are recorded in ice-fish, which live in almost ice-cold water (Rigby, 1968). We believe that during adaptation to higher body temperatures there are several independent evolutionary lines in collagen, each pointing toward a higher imino acid content. This is an example of a convergent evolution.

Eastoe long ago pointed out (Eastoe, 1957) that the number of hydroxy amino acid residues is constant in most collagens, but this fact has been disregarded in structural considerations and in model-building. This issue also has been clouded by the fact that collagen of certain worms has an exceptional amount of hydroxyl groups both as hydroxyproline and as serine and threonine (Maser and Rice, 1962). Rigby (1967) suggested that there was some inverse correlation between the thermal stability and serine content.

I AMINO ACID COMPOSITION

The findings mentioned above were confirmed by our studies on collagen extracted with acidic buffers and purified exhaustively by repeated precipitations. The charge profile of reconstituted SLS particles under the electron microscope, prepared by one of us (J.P.) in Prof. Kühn's laboratory, was similar throughout the investigated series (Nordwig and Hayduk, 1967). Bornstein (1968) suggests that variation is possible mostly in the non-helical regions, but their spacing seems to be constant.

We compiled a series of comparisons of amino acid compositions (Table I). The constancy of hydroxy-amino acids is confirmed, with the exception of the earthworm (*Lumbricus*). The increase of content of imino acids during the course of evolution is especially distinct in the homothermic species. The content of aliphatic hydroxy-amino acids, serine and threonine, complements the imino acids fairly closely. The amino acids mentioned: proline, serine and threonine, plus alanine, differ in the genetic code by the first letter only, so that single base mutations can produce these changes. These amino acids take approximately the same space in the secondary structure, and the position of the

TABLE I

Amino Acid Contents of Collagen from Various Animals

Group of amino acids	Man	Monkey	Pig	Chick	Frog	Flounder	Dogfish	Rayfish	Hagfish	Lamprey	*Loligo*	*Mytilus*[5]	*Lumbricus*	*Metridium*
Hydroxy amino acids[1]	148	136	154	148	148	152	146	170	165	152	159	163	290	180
Serine+threonine	49	47	53	46*	78	85	71	101	98	81	70	86	136	86
Imino acids	222	225	223	212*	167	166	171	145	154	173	150	147	164	137
Imino acids+Ser+Thr+Ala	374	382	387	385	367	371	361	343	368	375*	312	321	397	295
Nonpolar amino acids[2]	653	673	650	648	655	656	642	611	651	647*	614	597	509	563
Acidic amino acids[3]	115	104	118	120	117	105	114	114	109	116*	154	158	150	162
Basic amino acids[4]	88	91	86	87	83	90	101	108	78	89	93	85	51	107

[1]Hypro+Thr+Ser+Tyr+Hylys; [2]Pro+Gly+Ala+Val+Met+Ileu+Leu+Phe; [3]Asp+Glu; [4]Hylys+Orn+Lys+His+Arg;
[5]byssus-threads

*The asterisks indicate marked differences between the adjacent groups. The contents are expressed as residues/1000. The data are based on the amino acid analyses by Pikkarainen (1968) and Pikkarainen et al. (1968), except those on man, monkey and *Loligo* (unpublished data).

hydroxyl group is rather similar. Table II shows the same comparison extended to imino acids and thermal stability. The ratio of imino acids to the sum of serine and threonine is the most sensitive index of this evolutionary trend (Fig. 1). It would be expected from these conclusions that proline is hydroxylated preferentially in such positions of the peptide chain which carried the hydroxyl of serine or threonine at the earlier stages of development.

There is no difference between invertebrates and vertebrates in the hydroxy or imino acids but a characteristic dissimilarity can be located in the acidic amino acids (Table I). The amino acid composition suggests that the complementary amino acid would be alanine or, at least a nonpolar amino acid. Whether the amino acid compositions can be related to the capacity to form bone remains to be studied. Collagen of certain lower animals, for example the sea anemone (Metridium), contains abundant hydroxylysine, but generalizations cannot be made. The presence of cystein is common in invertebrate collagen.

We tried to assess the overall differences between the amino acid composition of collagen of various species. This was done by totaling the differences observed within every

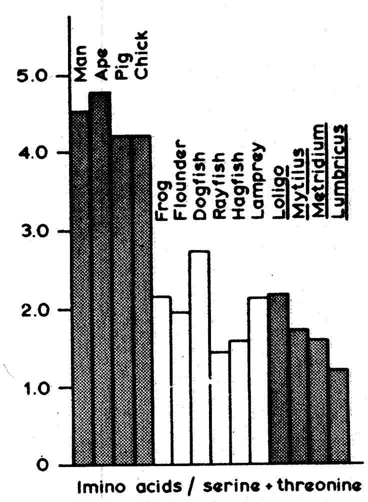

Fig. 1. Ratio of imino acids to serine+threonine in collagens from various species.

TABLE II

Relation of Hydroxy and Imino Acids to the Thermal Stability of Citrate-soluble Collagen

Origin of collagen	Hydroxyproline	Threonine	Serine	Total Hydroxy residues	Imino acid residues
Rayfish	61	30	71	170	145
Hagfish	62	27	71	165	154
Flounder	63	20	65	152	166
Lamprey	66	23	58	152	173
Lumbricus	151	52	84	290	164
Metridium	70	39	47	180	137
Dogfish	72	25	46	146	171
Loligo	70	26	44	159	150
Frog	65	22	56	148	167
Man	91	18	31	148	222
Chick	99	19	27	148	212

The origins of collagen are listed in order of the increasing T_D-values of collagen. Cf. legend of Table I.

pair of species on each amino acid, but disregarding the hydroxylation of proline and lysine. If the contents of individual amino acids are expressed as residues per thousand, the sum of the differences would be 2000 maximally, if all the amino acids were different. The figure is too small because it does not include the mutually cancelling differences in the amino acid sequence. Nevertheless, we think that it reflects the number of mutations when their total number is low. In many proteins this kind of comparison would be too superficial, but a similar approach has been tried with certain success (Metzger, Shapiro, Mosiman and Winton, 1968). With collagen this procedure seems to produce meaningful results (Table III). The order of amino acid composition follows the morphological taxonomy, as expected. Encouraged by this observation, we attempted to align the species on a plane (Fig. 2) on the basis of

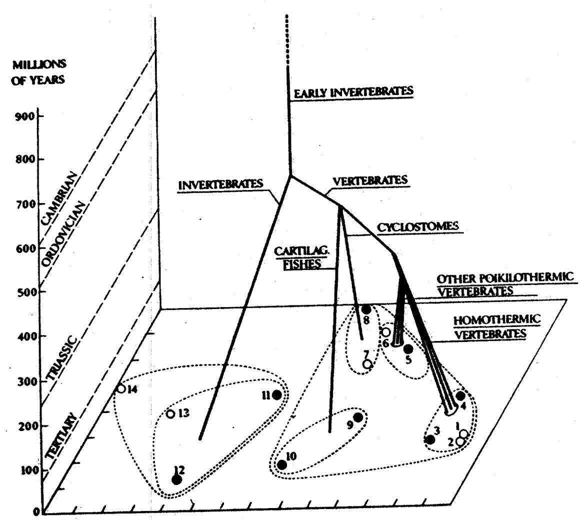

Fig. 2. Summary on the evolution of collagen. The differentiation of the α chains in homothermic and certain poikilothermic vertebrates is indicated. The location of the species on the horizontal plane is based on the differences in the amino acid compositions as explained in the text. Key to numbering: 1 man, 2 monkey (Cercopithecus), 3 pig, 4 chick, 5 frog, 6 flounder, 7 lamprey, 8 hagfish, 9 dogfish, 10 rayfish, 11 squid, 12 sea mussel (Mytilus), 13 sea anemone (Metridium), 14 earthworm (Lumbricus).

TABLE III

Differences in the Amino Acid Composition of Collagen from Various Species

(Cf. legend of Table I.)

Species	Man	Monkey	Pig	Chick	Frog	Flounder	Dogfish	Rayfish	Hagfish	Lamprey	Loligo	Mytilus*	Lumbricus	Metridium
Man	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Monkey	41	-	-	-	-	-	-	-	-	-	-	-	-	-
Pig	43	45	-	-	-	-	-	-	-	-	-	-	-	-
Chick	76	83	71	-	-	-	-	-	-	-	-	-	-	-
Frog	166	162	165	142	-	-	-	-	-	-	-	-	-	-
Flounder	200	191	195	170	62	-	-	-	-	-	-	-	-	-
Dogfish	142	146	127	114	116	134	-	-	-	-	-	-	-	-
Rayfish	196	220	205	214	180	182	108	-	-	-	-	-	-	-
Hagfish	240	233	241	226	104	72	170	170	-	-	-	-	-	-
Lamprey	166	169	165	138	54	60	94	166	96	-	-	-	-	-
Loligo	186	210	207	208	150	182	136	136	178	148	-	-	-	-
Mytilus*	247	273	246	273	234	282	197	163	276	250	151	-	-	-
Lumbricus	278	312	275	292	238	271	230	199	268	255	214	180	-	-
Metridium	276	294	269	282	259	290	196	181	297	266	137	132	213	-

*Byssus-threads

differences in amino acid composition, and assuming that these differences are proportional to the number of mutations in collagen. The differences had to be adjusted to fit into a network; this was done with an electronic computer by minimizing the sum of squares of the necessary changes expressed proportionally. In this way we obtained "coordinates" for 8 species (shown by black dots on the horizontal plane in Fig.2). The others were fitted by dividers with reference to the nearest computer-calculated points. It is the advantage of this presentation that more comprehensive evolutionary schemes could be constructed, with time as the third axis. Collagen of the earthworm (Lumbricus), occupies a special position on the "map" and it may not belong to this scheme at all but to another family of collagens. We also tried to include data available from the literature. Unfortunately, the results were not consistent, which partly is accounted for by analytical differences.

The same procedure as demonstrated in Table III was applied to specific groups of amino acids, for example, on aliphatic hydroxy-amino acids (Table IV) to demonstrate the gap between the homo- and poikilothermic animals. The difference between vertebrates and invertebrates is shown analogously on the basis of the acidic amino acids (Table V).

It is our belief, the evolution of collagen did not necessarily concur with the evolution of other chemical constituents or with the morphological features. Thus, classifications cannot be based on a single morphological characteristic or protein. For example, hagfish and lamprey differ in collagen more than expected from the morphological similarity. The evolutionary pattern of the various parameters is quite diffuse, especially within the group of bony fishes.

II CHAIN STRCUTURE

Analogous comparative calculations were applied to the various chains of tropocollagens, relying entirely on published values of the amino acid composition (Table IV). (Pikkarainan and Kulonen, 1969)

TABLE IV

Differences in the Contents of Serine+threonine in Collagens from Various Species

(Cf. legend of Table I)

Species	Man	Monkey	Pig	Chick	Frog	Flounder	Dogfish	Rayfish	Hagfish	Lamprey	*Loligo*	*Mytilus**	*Lumbricus*	*Metridium*
Man	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Monkey	2	-	-	-	-	-	-	-	-	-	-	-	-	-
Pig	4	6	-	-	-	-	-	-	-	-	-	-	-	-
Chick	3	1	7	-	-	-	-	-	-	-	-	-	-	-
Frog	29	31	25	32	-	-	-	-	-	-	-	-	-	-
Flounder	36	38	32	39	7	-	-	-	-	-	-	-	-	-
Dogfish	22	24	18	25	7	14	-	-	-	-	-	-	-	-
Rayfish	52	54	48	55	23	16	30	-	-	-	-	-	-	-
Hagfish	49	51	45	52	20	13	27	3	-	-	-	-	-	-
Lamprey	32	34	28	35	3	4	10	20	17	-	-	-	-	-
Loligo	21	23	17	24	8	15	1	31	28	11	-	-	-	-
*Mytilus**	37	39	33	40	8	1	15	16	13	5	16	-	-	-
Lumbricus	87	89	83	90	58	51	65	35	38	55	66	51	-	-
Metridium	37	39	33	40	8	1	15	16	13	5	16	0	51	-

*Byssus-threads.

TABLE V

<u>Differences in the Contents of Acidic Amino Acids in Collagens from Various Species</u>

(Cf. legend of Table I.)

Species	Man	Monkey	Pig	Chick	Frog	Flounder	Dogfish	Rayfish	Hagfish	Lamprey	Loligo	Mytilus*	Lumbricus	Metridium
Man	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Monkey	12	-	-	-	-	-	-	-	-	-	-	-	-	-
Pig	3	15	-	-	-	-	-	-	-	-	-	-	-	-
Chick	5	17	2	-	-	-	-	-	-	-	-	-	-	-
Frog	2	14	1	3	-	-	-	-	-	-	-	-	-	-
Flounder	10	2	13	15	12	-	-	-	-	-	-	-	-	-
Dogfish	1	11	4	6	3	9	-	-	-	-	-	-	-	-
Rayfish	1	11	4	6	3	9	0	-	-	-	-	-	-	-
Hagfish	6	6	9	11	8	4	5	5	-	-	-	-	-	-
Lamprey	1	13	2	4	1	11	2	2	7	-	-	-	-	-
Loligo	39	51	36	34	37	49	40	40	45	38	-	-	-	-
Mytilus*	43	55	40	38	41	53	44	44	49	42	4	-	-	-
Lumbricus	35	46	32	30	33	45	36	36	41	34	5	9	-	-
Metridium	47	58	44	42	45	57	48	48	53	46	8	4	12	-

*Byssus-threads.

TABLE VI

Differences in the Amino Acid Composition of the Different Chains in Collagens of Various Species

Species	$\alpha 1$ vs. $\alpha 2$	$\alpha 1$ vs. $\alpha 3$	$\alpha 2$ vs. $\alpha 3$
Calf	126	53	82
Chick	124	23	129
Cod	76	36	60
Species	$\alpha 1$ vs. $\alpha_{Myx.}$	$\alpha 2$ vs. $\alpha_{Myx.}$	$\alpha 3$ vs. $\alpha_{Myx.}$
Mammals	262	240	265
Birds	256	212	259
Bony fishes	146	119	104

Piez, Eigner and Lewis, 1963; Bornstein and Piez, 1964; Heidrich and Wynston, 1965; Piez, 1965; François and Glimcher, 1967a,b; Paz, Salazar and Gaite, 1967.
$\alpha_{Myx.} = \alpha$ chain of collagen from hagfish (Myxine glutinosa).

Even if the data of various authors differ considerably, we believe that some conclusions can be drawn. The chains developed divergently, but $\alpha 3$ chain resembles $\alpha 1$ more than $\alpha 2$ chain. The collagen of mammals and birds differs more from hagfish collagen than does collagen from bony fishes. During the course of evolution there was a tendency toward development of stable forms, as indicated by the increasing ratio of imino acids to serine+threonine, which is higher in the $\alpha 1$ and $\alpha 3$ chains than in the $\alpha 2$ chain. The $\alpha 1$ and $\alpha 3$ chains are more stable than $\alpha 2$ which is reflected also in their higher thermal stability (Hollmén and Kulonen, 1964) and in the slower metabolic turnover (Heikkinen and Kulonen, 1968).

This divergent development suggests a triple-helix composed of similar chains in the ancestral collagen analogous to the evolution of myoglobin and hemoglobin. Indeed, the results of gel electrophoresis and column chromatography on carboxymethyl cellulose suggest that the chains in the available soluble collagens from invertebrates—lamprey, hagfish, rayfish and dogfish—have only

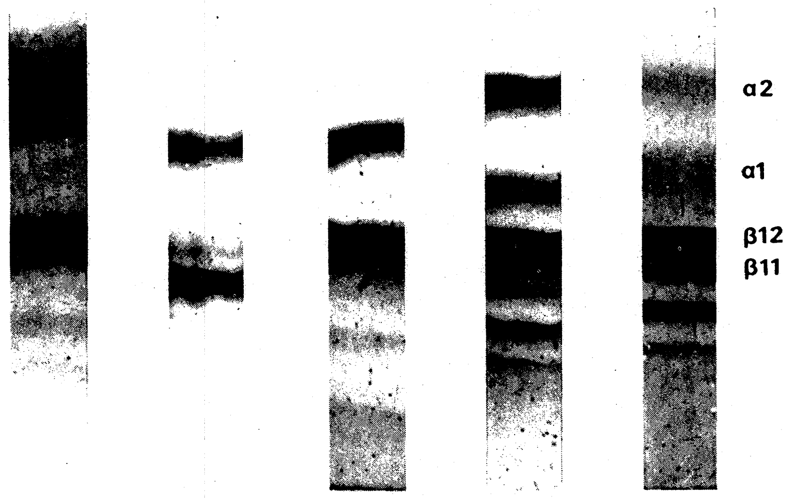

Fig. 3. Starch-gel electrophoretic patterns of denatured citrate-soluble body-wall collagens from various species.

one electrophoretically-distinguishable α component (Fig.3). The higher animals from reptiles upwards all have two or three α components with the corresponding β components. The collagens of bony fishes are variable in this respect. The analysis is obscured by the simple degradation of the fish collagens, which makes analysis of the components uncertain. In the gel electrophoresis pattern of lamprey collagen there is a faint band which could be interpreted as α1 chain. We must consider the possibility that minor amounts of another α chain are present.

Taking into account the knowledge of amino acids we summarised our ideas in Fig. 2, with some reservations about bony fishes.

III STABILITY AND CROSS-LINKING

The more highly developed vertebrates seem to have greater efficiency in forming cross-linked collagen that is resistant to solvents. The rate of the gelatinization of insoluble skin collagen of higher species is slower with minimal heating (Table VII) although we should be cautious in drawing conclusions because of the different metabolic

TABLE VII

Solubilization of Insoluble Vertebrate Skin Collagen

Species	Extracted* at		Residue
	+40°	+65-90°	
Cow	0.5	10.9	88.6
Calf	2.7	76.0	21.3
Guinea pig, adult	9.2	74.9	15.9
Guinea pig, growing	4.2	91.2	4.6
Chick	6.0	72.5	21.5
Snake	2.7	32.3	65.0
Toad	16.9	73.7	9.4
Burbot	69.3	24.1	6.6
Pike	78.4	17.7	3.9
Hagfish	55.9	26.9	17.2
Lamprey	89.9	7.5	2.6

*Into 0.01 M sodium acetate buffer, pH 4.8, in percent of total insoluble collagen (Pikkarainen, 1968).

turnovers of the fractions. Heikkinen (1968) has shown that the 80°C-soluble fraction of insoluble rat collagen contains associated substances different from those in the respective 40°C-soluble fraction. The differences between the species (Table VII) may depend on the associated substances. The existence of a biphasic temperature-shrinkage curve in fish collagens (Pikkarainen et al., 1966) indicates weaker crosslinks, which are hidden by the strong crosslinks in collagens of higher species. This may have a structural basis either in the cross-linking groups or in the accessory substances in the insoluble collagenous fibres. A good example on the effect of the noncollagenous substances is the byssus-threads of the sea mussel (*Mytilus*), which have the thermal shrinking point at about 90°C whereas the skeleton collagen of the same animal shrinks at about 55°C.

Nothing is known about the lysine cross-links during the course of the evolution.

IV DISCUSSION

Judging from the thermal stability and for that matter from the imino acid content, of collagen, there has been a similar but independent natural selection process in worms and in vertebrates. Qualities other than thermal shrinkage and imino acid content are quite different. In the development of the animals with high and constant body temperature the improvement of the thermal stability of collagen is one of the decisive factors, although there are numerous enzyme proteins which also denature just above the body temperature of mammals and birds. Preliminary work from our laboratory indicates that during fever an excess of hydroxyproline is excreted.

What was the original collagen? Presumably it contained every third residue of glycine in common with elastin and silk proteins (Rudall, 1968). Furthermore, there may have been one hydroxyl for every six residues, abundantly acidic and other polar amino acids, and enough imino acids to maintain the helix form even at temperatures near zero. We may speculate that there was originally a tripeptide Gly-X-Y, which was multiplied a thousand-fold; the resulting peptide chain was able to form the collagen-fold and the triple-helix.

What benefit does the organism derive from differentiation of the three chains? Veis, Anesey and Mussell (1967) have proposed a special role for the $\alpha 2$ chain in the formation of the fibrous structure. Is it too farfetched to speak of eventual intramolecular specialization between the chains (Table VIII) in determination of stability, in cross-linking, in intermolecular organization or in the bone formation?

The report given by Dr. Miller during this symposium on the single-chain collagen from the sternal cartilage of chicken is very exciting to us, because in our scheme the cartilaginous fishes are the most advanced species having single-chain collagen, whereas many bony fishes have differentiated chains. The present scheme (Fig. 2) applies to

TABLE VIII

Comparison of the Chains in Collagen

Thermal stability	$\alpha 1$, $\alpha 3 > \alpha 2$
Formation of quaternary structure	$\alpha 2$ decisive
Aldehyde content	$\alpha 1 = \alpha 2$
Sensitivity to β-APN	$\alpha 1 > \alpha 2$
Carbohydrate content	$\alpha 1$, $\alpha 3 > \alpha 2$
Content of bound phosphorus	$\alpha 2 > \alpha 1$, $\alpha 3$

Hollmên and Kulonen, 1964; Veis et al., 1967; Rojkind and Juarêz, 1966; Glimcher and Krane, 1968.

fibrous collagen. We suggest that the collagen of ectodermal origin (e.g. from basement membrane) may belong to a different system, along with worm-cuticle collagen.

ACKNOWLEDGEMENTS

Institutional support from the U.S. Department of Agriculture, Foreign Research and Development Division, and from Emil Aaltonen Foundation (J.P.) is gratefully acknowledged.

REFERENCES

Bornstein, P. (1968). Science 161, 592.

Bornstein, P. and Piez, K.A. (1964). J. Clin. Invest. 43, 1813.

Eastoe, J.E. (1957). Biochem. J. 65, 363.

Eastoe, J.E. (1968). In Treatise on Collagen, vol. I, p. 1. Ed. by Ramachandran, G.N.Academic Press, London.

Eastoe, J.E. and Leach, A.A. (1958). In Recent Advances in Gelatin and Glue Research, p. 173. Ed. by Stainsby, G. Pergamon Press, London.

François, C.J. and Glimcher, M.J. (1967a). Biochim. Biophys. Acta 133, 91.

François, C.J. and Glimcher, M.J. (1967b). Biochem. J. 102, 148.

Glimcher, M.J. and Krane, S.M. (1968). In Treatise on Collagen, vol. 2-B, p. 67. Ed. by Gould, B.S. Academic Press, London.

Gross, J. (1963). In Comparative Biochemistry. A Comprehensive Treatise, vol. 5, p. 307. Ed. by Florkin, M. and Mason, H.S. Academic Press, New York.

Gustavson, K.H. (1956). The Chemistry and Reactivity of Collagen. Academic Press, New York.

Harrington, W. F. and Rao, N. V. (1967). In Conformation of Biopolymers, vol. 2 , p. 513. Ed. by Ramachandran, G.N. Academic Press, London.

Heidrich, H.-G. and Wynston, L.K. (1965). Z. physiol. Chem. 342, 166.

Heikkinen, E. (1968). Acta Physiol. Scand. Suppl. 317.

Heikkinen, E. and Kulonen, E. (1968). Biochim. Biophys. Acta 160, 464.

von Hippel, P.H. (1967). In Treatise on Collagen, vol. 1, p. 253. Ed. by Ramachandran, G.N. Academic Press, London.

von Hippel, P.H. and Wong, K.-Y. (1963). Biochemistry 2, 1387.

Hollmên, T. and Kulonen, E. (1964). Biochim. Biophys. Acta 93, 655.

Josse, J. and Harrington, W.F. (1964). J. Mol. Biol. 9, 269.

Maser, M.D. and Rice, R.V. (1962). Biochim. Biophys. Acta 63, 255.

Metzger, H., Shapiro, M.B., Mosimann, J.E. and Winton, J.E. (1968). Nature 219, 1166.

Nordwig, A. (1968). Mitteilungen aus der Max-Planck-Gesellschaft, 5, 316.

Nordwig, A. and Hayduk, U. (1967). J. Mol. Biol. 26, 351.

Paz, M.A., Salazar, M.E. and Gaite, E.E. (1967). Arch. Biochem. Biophys. 119, 485.

Piez, K.A. (1965). Biochemistry 4, 2590.

Piez, K.A., Eigner, E. A. and Lewis, M.S. (1963). Biochemistry 2, 58.

Pikkarainen, J. (1968). Acta Physiol. Scand. Suppl. 309.

Pikkarainen, J. and Kulonen, E. (1965). Acta Chem. Scand. 19, 280.

Pikkarainen, J. and Kulonen E. (1967). Scand. J. Clin. Lab. Invest. 19, Suppl. 95, 40.

Pikkarainen, J. and Kulonen, E. (1969). Nature 223, 839.

Pikkarainen, J., Tantanen, J. and Kulonen, E. (1966). In Biochimie et Physiologie du Tissu Conjonctif, p. 303. Ed. by Compte, Ph. Lyon.

Pikkarainen, J. Rantanen, J., Vastamäki, M., Lampiaho, K., Kari, A. and Kulonen, E. (1968). Eru. J. Biochem. 4, 555.

Rigby, B.J. (1967). Nature 214, 87.

Rigby, B.J. (1968). Nature 219, 166.

Rojkind, M. and Juárez, H. (1966). Biochem.Biophys. Res. 25, 481.

Rudall, K.M. (1968). In Treatise on Collagen, vol. 2-A, p. 83. Ed. by Gould, B.S. Academic Press, London.

Veis, A. Anesey, J. and Mussell, S. (1967). Nature 215, 931.